Proceedings of the 34th Accra Bespoke Multidisciplinary Innovations Conference & the Africa AI Stakeholders Summit
Academic City University College, Accra Ghana
19th – 21st December, 2022
www.isteams.net/accrabespoke2022

Improved Privacy Protection Model for Prevention of Data Over-Collection in Smart Devices

Oketayo, Abimbola M. & Ojo, Adebola K.
[1,2]Department of Computer Science
University of Ibadan
Ibadan, Nigeria
E-mails: oketeeabimbola@gmail.com; adebolak.ojo@gmail.com
Phones: [1]+2348038353949,[2]+2347032736013

ABSTRACT

In this study, an attempt was made using machine learning algorithm with the user data store in the mobile cloud framework to solve the problem of data over-collection. This was achieved by designing a model using the security risk level of the applications and the corresponding class level of the users on the smartphone that will help in preventing smartphone apps from accessing and collecting users' private data while still within the permission scope. Users can store information in the cloud environment where the huge numbers of users are involved. We develop a mobile agent simulator to generate data, and determine the security risk level of the apps on users' data with the class level of the data. The permission model was designed to determine whether the app is granted permission to access user's data or not. The data was trained with the use of Neural Network. The evaluation metrics used were accuracy and comparison. For accuracy, the algorithm was compared with the existing algorithm. The data analysis showed that there was restriction for apps accessing the users' data. The model if deployed on the smartphone will prevent apps from over collect users' data even while still within the permission scope. This study proved that neural network with mobile cloud computing can be applied to prevent data over-collection in smart devices.

Keywords: Data over-collection, Private data, Smartphone, Security risk, Class level, Simulator, Privacy

Proceedings Reference Format
Oketayo, A.M. & Ojo, A.K. (2022): Improved Privacy Protection Model for Prevention of Data Over-Collection in Smart Devices. Proceedings of the 34th Accra Bespoke Multidisciplinary Innovations Conference & The Africa AI Stakeholders Summit. Academic City University College, Accra Ghana, 2022. Pp 137-144. www.isteams.net/accrabespoke2022
https://doi.org/10.22624/AIMS/ACCRABESPOKE2022/V34P12

1. INTRODUCTION

Data over-collection is a way by which smartphone apps can access and collect more data from users' smartphones than what is expected within the permitted scope (Yibin L, Wenyun D, Zhong M,, 2016). Smartphones are the most commonly used electronic devices now-a-days because of their sizes and functionality, the usage of smartphones are enormously high since users store information that are personal and sensitive data such as passwords, e-mails, photos, videos, contacts, and so on their devices.

Hence, security is seeming to a threatening factor, for this reason, the security and privacy of smartphone data has become a major concern in the Internet of Thing (IoT) world. Aside from the usefulness and functionality, applications on smartphone pose significant security risks. Phishing, web browser exploits, add-ons such as formatting flaws, scripting issues, security by-passes, protocol handling, and downloading executable files are all examples of web-based risks. All offline risks, such as software tampering, authentication, authorization, configuration management, sensitive information, session management, and cryptography, are included in application-based threats (Yibin L, Wenyun D, Zhong M,, 2016). These threats mainly focus on retrieving the information of the users from their devices.

There is no 100% of leakage and blocking of the data collection in the smart devices, but there is need to prevent, detect, and provide solution to the problem of data over-collection. With the rapid advancement of database, networking, and computer technologies, as well as the Internet of Things (IoT), there is need to protect private or sensitive data from being accessed by applications. Data can be digitally integrated and analyzed, which has contributed to the development of data on the one hand, but on the other hand, easy access to private data poses a threat to individual privacy (Alpaydin, 2005). In this paper, we present models and algorithms for privacy protection users' data to prevent data over-collection.by apps (Yibin L, Wenyun D, Zhong M,, 2016)

2. RELATED WORKS

Privacy, according to (Solove, 2008), defined privacy as a term that encompasses a broad and heterogeneous range of related things. According to Privacy International, privacy is a multi-dimensional notion that includes four components: the body, communications, territory, and information. The term "bodily privacy" refers to people's physical protection from danger or external threats, whereas "personal privacy" refers to the privacy of their information. On the security of information transmitted between two parties via any channel, such as mail, letters, and telephone and Setting boundaries or limits on physical space or property, such as the home, workplace, or public locations, is what territorial privacy is all about. The term "privacy protection" refers to a set of tactics and procedures for ensuring the privacy, availability, and integrity of data (Alexander A, Varfolomeev 1, Liwa H and Zahraa C,, 2020).

This study is a supervised algorithm since it builds a mathematical model of a data set that contains both inputs and desired outputs. Since we are preventing access to users' data with a security risk label and class level, to determine the security class level of such apps on users' data and classification is being used. As a result of this, it can be applied without prior knowledge of structure to be discovered. (Anand V.R.S, Janani E.S.V. , 2017) developed a strategy to tackle the problem of data over-collection in smart devices by combining a mobile cloud framework with a key policy attributes-based encryption (KP& BE) model. Information is kept in the cloud in his model, and any data must be confirmed by the user before it can be collected. The user is aware of the information that will be gathered. The approach has a flaw in that it cannot successfully identify sensitive or private data of users by analyzing the flow of data in apps together.

(Rashmi C, and Rajabhushan C. , 2018) worked on privacy protection of for preventing data over-collection in smart cities, they noted that because smartphones are unable to manage users' sensitive data, there may be privacy leakage as a result of over-collection of data which is the major setback in smart cities. The services to manage users' data and enable fine-grained access control with encryption or decryption are designed into a mobile cloud architecture. They came to conclusion that because smartphones largely deal with basic app operations and key activities take place in the cloud, the only thing the strategy could effectively ensure was storage space.

3. METHODS AND PROCEDURE

The mobile agent simulator generated Dataset, the dataset generated was used to simulate usage data, these datasets consist of data accessed by different apps, and they are a set of structured data with labels of security risk levels, security class, security level, probability and permission for each app scenario for each server interactions.

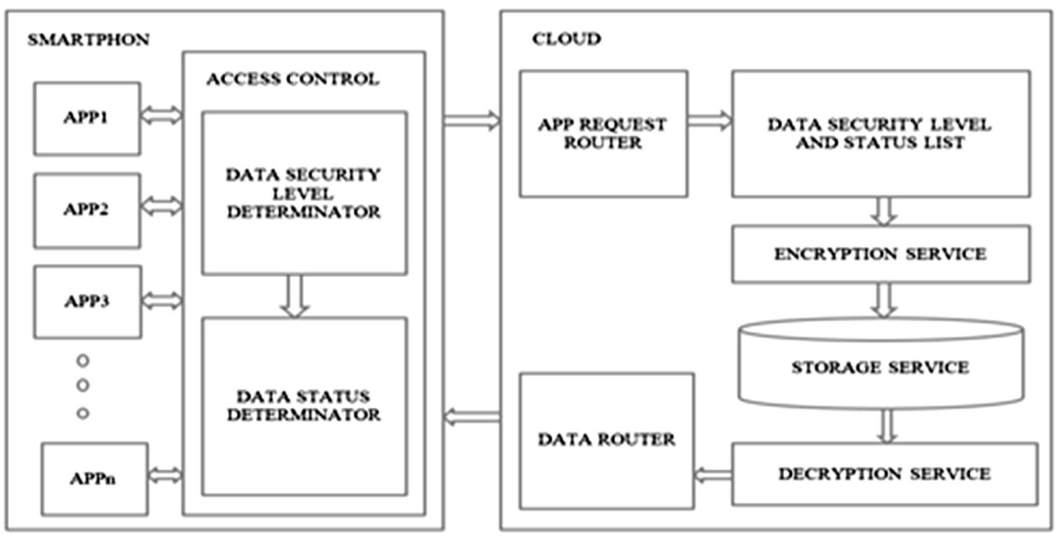

Figure 1: Proposed Methodology

Figure 1 shows the proposed model designed to provide solution to the problem of users' data over-collection. The model is designed to reduce the over-collection of data by smartphone applications that have acquired lawful permission to collects users' sensitive or private data. The model structure is made up of two major entities that communicate with one another in a specific way to share information. These two entities include: Smartphones and the Cloud. The Smartphone entity contains the access control module that determines the security risk level of data and the different status of data in the security risk class, whereas the Cloud comprises of The App request router, Data router module, Data security leve and status list module, Encryption and decryption module and, Storage service module.

Permission model for over-collection of users' data security

Currently, smartphone operating systems only provide fine coarse-grained permission authorization, which is simply all for nothing. This is the main source of data ove-collection in smartphones. We add different permission authorizations to the smartphone's security and consumption model to quantify the influence of coarse-grained permission authorisation towards data with varied security levels. We set several distinct sorts of permission authorizations, which are to all, specific, and none, according to the cloud storage service. Furthermore, the proportion of data accessing can be used to define the permission to specific. We denoted the permission of an app as N/M if it has authorization to access the data of N users, if it has M data in total. We consider security level (SL) and class level (CL) to characterize the impact of authorization for different security level data comprising distinct classes of data.

We multiply N/M by SL/CL to get the final authorization, which is

$$perm = \frac{SL}{CL} \times \frac{N}{M} \qquad \text{..... (1)}$$

Algorithm for assign permission to user data

Inputs: appID, userPassID, user data to access n, data class t, default security risk DSR t
Output: The ability to access p is returned as an output.
1: Use the appID, userPassID, n, and t parameters to send an Access Control Service request;
2: SR DSR is the user's default security level.
3: Determine the security level of this app, sl.
4: Determine the security level of this app (class cl);
5: Determine the total size of the t type data m
6: Calculate the proportion of data n towards total data m, $perm = \frac{SL}{CL} \times \frac{N}{M}$*;*
7: Determine the data n's Security risk,

$$SR_i^d = \frac{SL^d}{CL^d} \times \left(1 - e^{-\gamma * N_i^d / M_i}\right);$$

8: if $(sl/cl) > (SL/CL)$ *then*
9: return no permission;
10: else
11: Calculate the security risk of app appID,

$$SR_i^U = \sum_{d=0}^{m} \frac{SL^d}{CL^d} * (1 - e^{-\gamma * N_i^d / M_i});$$

12: if slappID > SL then
13: return no permission;
14: else
15: return permission;
16: end if
17: end if

3.1 Risk Model for Over-Collection of users' data by app

To formulate the problem that may be caused by application accessing too much data, it is necessary to introduce the security risk model. Since data over-collection is a type of security risk that arises as a direct result of the likelihood of security violations and the failure of security protocols.

As a result, we model an application's security risk (SR) towards data d as follows:

$$SR_i^d = \frac{SL^d}{CL^d} \times Pro_i^d \quad \text{.....(2)}$$

Where SR_i^d denotes the security risk of an application *i*, over-collects the data *d*. Meanwhile, SL^d is the data *d* security level, CL^d is the class level of data *d* within the security level, and Pro_i^d is the likelihood of the app *i* using the data *d* to cause some security-related damage to the users' data, which may be expressed as:

$$Pro_i^d = 1 - e^{-\gamma * N_i^d / M_i} \quad \text{..... (3)}$$

´Y is the security risk coefficient of the application i's over-collecting data d's behavior, which can be adjusted by different apps and data, but fixed on a particular scenario. An unchecked data over-collection by applications running on the user's smartphone increases the likelihood of a security breach, leads to higher security risk. Combining Equations (2) and (3) above, we can use the amount of over-collected data (N/M) to formulate the security risk of apps *i* towards data *d* as:

$$SR_i^d = \frac{SL^d}{CL^d} \times (1 - e^{-\gamma * N_i^d / M_i}) \quad \text{..... (4)}$$

It is obvious that N_i^d / M_i ranges from 0 to 1 as probability function can only range from 0 to 2, where 0 means the application *i* has not acquired necessary permission to access the user data *d*, and 1 represents all data in a particular security level and security class *d* is over-collected by the application *i* . We compute the relation between the security risk and the amount of over-collected data based on different security levels. In actuality, many apps acquire an excessive amount of data from multiple sources. We define the security risk posed by app *i* to a user (U) as follows:

$$SR_i^U = \sum_{d=0}^{m} \frac{SL^d}{CL^d} * (1 - e^{-\gamma * N_i^d / M_i}) \quad \text{..... (5)}$$

Finally, the security risk posed by a smartphone belonging to user U and containing A number of applications can be expressed as follows:

$$SR_i^U = \sum_{i=0}^{A} * \sum_{d=0}^{m} \frac{SL^d}{CL^d} * (1 - e^{-\gamma * N_i^d / M_i}) \quad \text{..... (6)}$$

Algorithm for modeling data upload to the cloud

Input: appID, userPassID, userData, DataClass.
Output: Upload Data to the cloud

Determine the request type made by appID, return T;
if (T == hardware) then
Assign permission to the app using the appID;
else
Send a request to Access Control Service, including appID, userPassID, userData, and DataClass; Access Control Service examines the app's request parameters and accessControlList to determine whether or not this app has authorization.
return P;
if (P == true) then
Encryption Service encrypts data;
Save data to Cloud Storage using the labels appID, userPassID, userData and userDataClass;
else
return;
end if
end if

Algorithm for app request access verified

Input: appID, userPassID, userDataClass preview information of requesting data PD.
Output: the requested data's specific content D

Send request to cloud request router including appID, userPassID, userDataClass and PD;
the request router interprets the request and sends the request to the security level and status list for confirmation, get the result P;
if (P == true) then
request status list sends request to cryptography service for encryption;
cryptography service (after encryption) sends request with appID, userPassID, userDataClass and PD to Storage Service;
Storage Service finds the encrypted data by userID, userDataClass and PD;
Storage Service sends data with appID and userID to Decryption Service;
Decryption Service verifies permission authorization again by matching appID with data and return P1;
if (P1 == true) then
Decryption Service decrypts data to data router D;
return D;
else
return none;
end if
else
return

4. DISCUSSIONS

In this section, the results from the study were discussed. Table 1 presents the formal usages of four smartphones used for the experiment while Table 2 shows the results of our Mobile app evaluated based on Security Risk which shows the experimental analysis see Figure 3.

Table 1: Formal Usages of four smartphones

Device	Photo	Music	Movie&Video
Smartphone P	2.0MB	99MB	0MB
Smartphone Q	1.7MB	2MB	32MB
Smartphone R	1.6MB	5.9MB	27MB
Smartphone X	0.9MB	3.1MB	30MB

Table 2: Result of Mobile apps evaluated based on Security Risk (Our Result)

SECURITY	ID	LOCATION	PHOTO	CONTACT	U & P
LEVEL & STATUS	(3,(1,2,3))	(3,(1,2,3))	(2,(1,2,3))	(1,(1,2,3))	(2,(1,2,3))

Table 3: The average result of our model

Mobile Phone	Security Risk - Without Model	Security Risk - With Our Model
SMARTPHONE P	41.55	< 22.55
SMARTPHONE Q	38.08	< 20.22
SMARTPHONE R	38.90	< 26.97
SMARTPHONE S	41.50	< 21.55

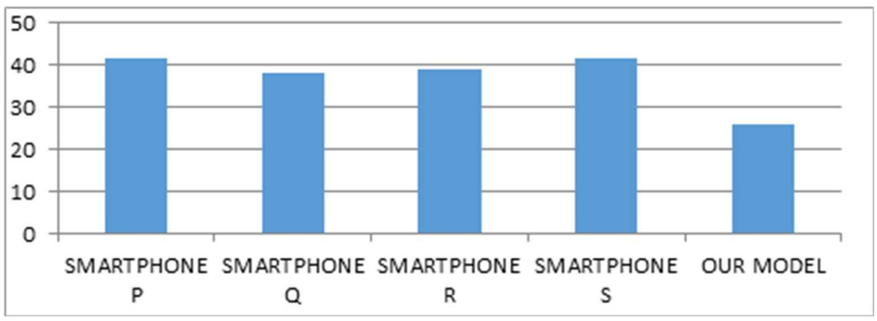

Figure 3: Graphical representation of our Model Result

4. RESULTS AND DISCUSSIONS

table 3 shows the results from the experiment, where four smart devices with five experimental applications tested separately in two different environments used for the evaluation. The security risks of these four smartphones with installed five apps are 41.55, 38.08, 38.90, and 41.50 in original environment unlike the available model in mobile-cloud framework environment proposed by (Yibin L, Wenyun D, Zhong M,, 2016), the security risk is less than

25.95. Using the same values for the apps in the original environment, our proposed model generated security risks values of 22.35, 20.23, 26.87, and 21.56 for the four (4) smartphones and an average of 22.75 calculated. The security dangers of smartphones in our model are obviously considerably lower than in the original environment.

Our solution model employed Application-Based Access Control (ABAC) model, which consists of three basic elements: application, roles, and permissions. The application is the subject that needs to interact with the data object, roles define the application's level of importance, and permissions define what can be viewed by a specific application. Roles and permissions are associated with each of the different applications (permission is a pair of objects and operations).

As such, a role is used to associate applications and permissions. The role of an application is the main object of the system, as it will serve as a link between the application and set of some permissions that is applicable to the application. An application in this model is the user's application that is capable of accessing users' data. A role represent what security level an app can access. Permission is an approval given to a particular operation to perform on a security class of a particular security level. This was responsible for the improvement in the results after the whole experiment.

5. CONCLUSION

The use of the proposed model designed for privacy protection for prevention of smartphone apps from accessing and collecting users' sensitive data has various effects, usefulness and its benefits cannot be overemphasized. The key to a successful privacy protection is the adoption of machine learning algorithm to carry out the work.

REFERENCES

1. Alexander A, Varfolomeev 1, Liwa H and Zahraa C,. (2020, November). Overview of Five Techniques Used for Security and Privacy Insurance in Smart Cities . *International Journal of Physics: Conference Series*. doi:10.1088/1742-6596/1897/1/012028
2. Alpaydin, E. (2005). *Introduction to Machine Learning.* MIT Press.
3. Anand V.R.S, Janani E.S.V. . (2017, May). Prevention of Data Over-Collection in Smart Devices. . *International Journal of Scientific & Engineering Research, 8*(5).
4. Rashmi C, and Rajabhushan C. . (2018). Privacy Protection for preventing Data over-collection in Smart City. *International Journal Of Innovative Research In Management, Engineering And Technology , 3*(4).
5. Solove, D. (2008). *Understanding Privacy.* Harvard University Press.
6. Yibin L, Wenyun D, Zhong M,. (2016). Privacy Protection for Preventing Data Over-Collection in Smart City. *IEEE Transactions on Computers, 65*(5), 1339-1350. doi:10.1109/TC.2015.2470247